# REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-02-

0298



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| August 27, 2002 | Final Performance Report | April 1999 - April 2002 |

**4. TITLE AND SUBTITLE**

Damage Assessment and Recovery from Information Warfare Attacks

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
F49620-99-1-0235

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Brajendra Panda

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

University of North Dakota

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Air Force Office of
Scientific Research

**10. SPONSOR/MONITOR'S ACRONYM(S)**
AFOSR

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for public release;
distribution unlimited

20021015 068

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Sensors at different Air Force operation sites collect information on various system parameters and send to the Air Force Computer Emergency Response Team (AFCERT) for analysis. Due to the massive amount and complex nature of data involved, this process, however, is inefficient and time consuming. It is rather desirable that each site pre-processes the data before transmitting to the AFCERT. For efficient processing of data at both local and global sites, development of a suitable format for storing data locally, and determining characteristics desired at the global site for the fusion of data obtained from different sites are important. In this research, the following issues have been addressed: 1) reduction of collected information for the diagnosis of attack, 2) efficient analysis of resultant data, 3) fast and accurate damage assessment, and 4) real-time recovery of the system.

**15. SUBJECT TERMS**

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | 6 | 19b. TELEPHONE NUMBER *(include area code)* |

# Final Report

**Proposal Title: Damage Assessment and Recovery from Information Warfare Attacks**

**Grant Number: F49620-99-1-0235**

**Principal Investigator: Dr. Thomas Wiggen**
**Compuer Science Department**
**University of North Dakota**
**Grand Forks, ND 58202**

(Initial PI was Dr. Brajendra Panda who moved to University of Arkansas in August 2001.)

***Major Accomplishments:*** The following key research activities have been accomplished as a result of this project.

1. Log segmentation models using (a) dependencies among data items, (b) fixed number of transactions, (c) fixed time window, and (d) fixed file size have been developed. Each of these methods organizes the database log file for faster access during damage assessment.
2. A formal approach to data dependency method has been established. A concept called Directed Damage Demonstration Graph has been developed which aids in faster and more accurate damage assessment.
3. For faster recovery a transaction fusion technique has been developed. This method combines several transactions and executes them as one transaction during recovery.
4. A data reduction technique has been developed which uses simple and compact data structures for faster damage assessment. Any of these structures can be used during damage assessment instead of the scanning the huge log file.
5. A data classification approach has been designed in order to manage transactions in the database in a secure manner. An appropriate access control mechanism has been developed.
6. Logging transactions' activities using their semantics is necessary for accomplishing complete and accurate recovery of a damaged system. Appropriate models have been developed in this project.

***Executive Summary***

In order to expedite damage assessment techniques have been developed for segmenting the log into multiple smaller files. During recovery segments containing damaged data items are accessed and those data items are recovered. The first method for log segmentation uses the data dependency relationships and stores related transaction operations in one segment. In this effort, the main objective was to segregate the related data items with accuracy. An algorithm to cluster the log by grouping the related operations based on data dependency has been developed. Specific measures have been heeded so that only the affected data operations would be considered during recovery once the attacking transaction is identified. This facilitates skipping those parts of the log that are unaffected during damage assessment and recovery. An algorithm for recovery using the clusters has also been offered. Recovery and damage assessment by this model demonstrated outstanding performance gain over the traditional log approach. Various concepts have been presented that would enhance the damage assessment process. During

recovery one of the major concern is denial of service. Making the database available is equally important as keeping it consistent. The concepts such as critical links and cliques would help in carrying out faster damage assessment and making the data items available at the earliest. These concepts have been developed in a generic manner, hence can be used along with other recovery models where the damage assessment during recovery is quite time consuming. These concepts must be applied to the database at the time of the design. The granularity of the nodes defined in those concepts may be changed in accordance to the project pursuit.

One of the problems with data dependency based log segmentation is that it uses considerable computation in determining the dependency relationships, thereby, slowing down transaction execution. Although the method significantly accelerates damage assessment process, in general, the payoff would not be substantial unless attack detections are frequent. In order to alleviate these problems, focus was given on how to segment the log based on criteria that do not use much of the system time but yet, serve the purpose of quick damage assessment. In the first model, segmentation of the log based on number of committed transactions has been proposed. A segment, which is called *tuft,* stored a fixed number of committed transactions in this case. In the second approach, the tuft was built based on a fixed time window. All transactions that commit within a time window are added to a tuft in this situation In the third approach, a constant tuft size was maintained and operations of committed transactions that fit into the tuft were recorded. It was assumed that the tuft size was bigger than the largest transaction. In this method transactions were not allowed to span through multiple segments though it would have saved disk space in order to simplify the process of damage assessment. Algorithms have been developed for damage assessment that could be used with a log that is segmented based on any of the three approaches. The damage assessment algorithm generates a list of all the malicious and affected transactions. Using this information, any of the previously proposed approaches can be used to carry out the recovery process. Through simulation it has been proved that damage assessment process with a segmented log is definitely faster than the traditional approach where the log is not segmented. Damage assessment on a log segmented based on the size of the tufts is quicker because the number of bytes read from the log to find the affected transactions is much less. It did not take as many bytes as the damage assessment based on number of committed transactions nor did it show random behavior as in the case of damage assessment based on a time window for the tufts. The tufts based on all three methods were built using the same log and hence the damage assessment process could be compared with each other.

Necessary theory and concepts have been developed to make the data dependency approach more robust and general. These include classifications of read and write operations, a new definition of transaction and a new representation of the scheduler. The proposed scheduler stores more pertinent information than the conventional scheduler. Based on the theoretical support, appropriate damage assessment and recovery algorithm has been designed. This algorithm considers dependencies among data items accessed by various transactions to precisely identify affected data items in a damaged database and restores them to their consistent values.

Recovery is one of the main phases in defensive information warfare, and must be carried out in the shortest time possible to minimize denial of service. The recovery process involves undoing of malicious and affected transactions and redoing of affected transactions. A model

that fuses each set of malicious transactions or affected transactions occurring in groups into a single fused transaction has been designed. These fused malicious and affected transactions are undone in the undo process and then the fused affected transactions are re-executed in the redo process. As the number of transactions and total number of operations are minimized, executing these new sets of fused transactions during recovery expedites the process. An algorithm based on the concept of transaction fusion has been developed. The method aimed at recovering the system affected by a malicious activity. A simulation model was constructed to evaluate the performance of the transaction fusion model. The results from the simulation model showed that transaction fusion model always performed better than the traditional approaches.

Further more, several data structures have been developed to store transaction relationships in various formats. During damage assessment, any of these structures can be used to detect the set of affected transactions without accessing the log. Efficiency of these auxiliary structures has been tested through simulation. Four algorithms have been developed, each of which stores dependency relationships among transactions in a unique list format. In a post-intrusion detection phase, the stored dependency list, which is extremely compact in size, is accessed for damage assessment. As a result, the damage assessment process becomes much faster using the developed model compared to the methods that uses traditional log with huge amounts of data. Consequently, the unaffected portion of the database can be made available to the users quickly. A simulation of the model was developed, which proved the efficiency of the model. This model could be applied to the dependency based logging approach without any significant change.

A mechanism for classifying data into rigid and regular categories has been developed. Using this model, a regular user can read both the rigid and regular category data sets but could only write in the regular category data. Only a very small group of people, people such as database administrators, for example, has write access to data in the rigid category. The most obvious reason for developing this protection mechanism was to prevent intentional violation of an access restriction by a user. It is essential to ensure that each active transaction uses data only in ways consistent with the stated policy for use of the data. A user is allowed to exercise the right on a data object based on the right (s)he has on that object. To facilitate that scheme a protection domain has been defined and suitable access matrix has been developed. When a user executes a transaction, the protocol performs some simple boolean operations to determine whether to accept or to reject the transaction. This classification model helps contain the damage to a limited area of database in case of an attack, thus, improving the efficiency of damage assessment and recovery process.

Traditional logs are inadequate for recovery from information attacks since they do not contain pertinent information. A semantically rich logging protocol that records all necessary information required for the complete repair of databases that suffered from malicious attacks has been devised. Based on this log, a mechanism for recovery from system failures has been developed. In addition, two methods are proposed for complete damage assessment and recovery of an attacked database. The first method performs damage assessment and recovery concurrently, while the second method performs these tasks sequentially. Both mechanisms work very efficiently by re-executing the damaged parts of an affected transaction rather than the entire transaction. The unaffected parts of transactions are carefully identified and avoided. The

first method requires that the entire system enter a quiescent state in which no new transaction is processed until the recovery is complete. The second method creates lists of affected data items and affected transaction parts in the damage assessment phase and releases the unaffected part of the database for regular operations. This makes the system available to users while the recovery process continued. All protocols and methods developed in this research have been substantiated with appropriate algorithms.

***Personnel Supported:***

Faculty:

1. Brajendra Panda (while at the University of North Dakota)

Graduate Students:

2. Kazi A. Haque
3. Rajesh Yalamanchili
4. Prahalad Ragothaman
5. Rumman Sobhan
6. Chandana Lala

***Resulting Theses***

1. "Theory and Concept for Reconstructing Database by Considering Dependencies Among Data Items", Kazi A. Haque, M.S., Computer Science, July 2002.
2. "Transaction Fusion: A Novel Recovery Paradigm", Rajesh Yalamanchili, M.S., Computer Science, July 2001.
3. "Grouping Committed Transactions Based on Number, Space, and Time", Prahalad Ragothaman, M.S., Computer Science, July 2001.
4. "Development of a Semantically Rich Logging Protocol for Complete Damage Assessment and Repair of Databases", Rumman Sobhan, M.S., Computer Science, December 2000.
5. "Design and Analysis of a Data Structure for Establishing Transaction Dependencies", Chandana Lala, M.S., Computer Science, May 2000.
6. "Efficient Damage Assessment and Recovery Using Log Clustering", Sani Tripathy, M.S., Computer Science, May 2000.

These theses can be found at the University of North Dakota library.

***Resulting Publications:***

Ragothaman, Prahalad and Panda, Brajendra, "Modeling and Analyzing Transaction Logging Protocols for Effective Damage Assessment", In Proceedings of the 16th Annual IFIP WG 11.3 Working Conference on Data and Application Security, King's College, University of Cambridge, UK, July 29-31, 2002.

Sobhan, Rumman and Panda, Brajendra, "Sequential Damage Assessment and Recovery Using Semantic Logging", In Proceedings of the 2002 IEEE Workshop on Information Assurance, United States Military Academy, West Point, NY June 2002.

Ragothaman, Prahalad and Panda, Brajendra, "Dividing Database Log into Equal Size Segments for Efficiency", In Proceedings of the 17th International Conference on Computers and Their Applications, San Fransisco, California, April 4-6, 2002.

Panda, Brajendra and Haque, Kazi, "Extended Data Dependency Approach: A Robust Way of Rebuilding Database", In Proceedings of the 17th ACM Symposium on Applied Computing, Special Track on Database Systems, Madrid, Spain, March 10-14, 2002.

Sobhan, Rumman and Panda, Brajendra, "Reorganization of Database Log for Information Warfare Data Recovery", Database and Application Security XV, M. Olivier and D. Spooner (editors), Kluwer Academic Press, 2001.

Panda, Brajendra and Ragothaman, Prahalad, "Alternate Methods of Storing Committed Transactions in the Log for Their Future Re-execution", In Proceedings of the 5th World Multi-conference on Systemics, Cybernetics, and Informatics, Orlando, FL, July 22-25, 2001.

Lala, Chandana and Panda, Brajendra, "Evaluating Damage from Cyber Attacks: A Model and Analysis", IEEE Transactions on Systems, Man, and Cybernetics, Part A, Special Issue on Information Assurance, Vol. 31, No. 4, July 2001.

Sobhan, Rumman and Panda, Brajendra, "Reorganization of Database Log for Information Warfare Data Recovery", In Proceedings of the 15th Annual IFIP WG 11.3 Working Conference on Database and Application Security, Niagara on the Lake, Ontario, Canada, July 15-18, 2001.

Tripathy, Sani and Panda, Brajendra, "Post-Intrusion Recovery Using Data Dependency Approach", In Proceedings of the 2nd Annual IEEE Systems, Man, and Cybernetics Information Assurance Workshop, West Point, NY, June 5-6, 2001.

Panda, Brajendra, and Yalamanchili, Rajesh, "Transaction Fusion in the Wake of Information Warfare", In Proceedings of the 2001 ACM Symposium on Applied Computing, Special Track on Database Systems, Las Vegas, Nevada, March 11-14, 2001.

Lala, Chandana and Panda, Brajendra, "On Achieving Fast Damage Appraisal in case of Cyber Attacks", In Proceedings of the IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop, West Point, NY, June 6-7, 2000.

Lala, Chandana, Panda, Brajendra, and Sobhan, Rumman, "Storing Transaction Dependency Graphs for Damage Appraisal Following an Information Attack", In Proceedings of the 15th International Conference on Computers and Their Applications, New Orleans, LA, March 29-31, 2000.

Panda, Brajendra and Tripathy, Sani, "Data Dependency Based Logging for Defensive Information Warfare", In Proceedings of the 2000 ACM Symposium on Applied Computing, Special Track on Database Systems, Como, Italy, March 19-21, 2000.